

---

## Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** The fetal exposome and DNA methylation in cord blood

**Creator:** Sophia Harlid

**Principal Investigator:** Sophia Harlid

**Data Manager:** Sophia Harlid

**Affiliation:** Umeå University

**Funder:** FORMAS

**Template:** Swedish Research Council Template

**ORCID iD:** 0000-0001-8540-6891

### Project abstract:

An increasing number of children are afflicted by allergies and common signs, such as food sensitivity and eczema, often manifest during the first years of life. The biological mechanisms responsible for allergy development are still unclear but the intrauterine environment is of high importance. One explanation could be that reprogramming of epigenetic marks in fetal DNA (e.g. DNA methylation patterns) increases the risk of asthma and allergy. The overall aim of this study is to build a predictive model for identification of children at high risk of asthma and allergy. Taking advantage of NorthPop - a recently started prospective birth cohort in Northern Sweden - this project will follow 1000 mother-child pairs from pregnancy to 3 years of age. Samples and information from the mothers are collected during pregnancy and cord blood is collected from the newborns. To identify information about the mothers living situation, digital questionnaires are used and environmental pollutants are measured in blood and urine. DNA methylation in cord blood is then connected to lifestyle data. Finally we will use advanced statistical methods to assess if measured differences in DNA methylation can predict allergic risk. This project is uniquely designed to evaluate the impact of the fetal exposome on DNA methylation patterns. The extensive follow up makes it possible to identify mechanisms behind the increase in asthma and allergy and aid in developing predictive strategies for clinical use.

**ID:** 55164

**Start date:** 01-01-2020

**End date:** 31-12-2025

**Last modified:** 14-02-2025

**Grant number / URL:** 2021-01098

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# The fetal exposome and DNA methylation in cord blood

---

## General Information

### Project Title

The fetal exposome and DNA methylation in cord blood – early prediction of childhood asthma and allergy?  
– A study in the prospective birth cohort NorthPop

### Project Leader

Sophia Harlid

### Registration number at the Swedish Research Council

2019-01187

Formas: 2021-01098

### Version

2.0

### Date

2024-02-14

## Description of data - reuse of existing data and/or production of new data

### How will data be collected, created or reused?

The project currently (fall 2020) have three ongoing, or planned, data collections that are described below. More collections may be added to later versions.

#### NorthPop QA data

All NorthPop participants receive questionnaires regarding their lifestyle, health, the health of their child and so on, at regular intervals from the time they join NorthPop until they (either) withdraw or until the child is seven years old. This data collection is managed by NorthPop (Data manager: Richard Lundberg, [richard.lundberg@umu.se](mailto:richard.lundberg@umu.se)) and stored on their secure server. For this project, variables of interest will be requested from the NorthPop database when needed. Data extracted from the NorthPop database will be stored on an external hard drive (locked in safe when not in use) or on a secure server at ITS Umeå. Only personnel directly involved in the project will have access. All data will be returned to NorthPop for long term storage after project completion.

#### DNA methylation Data

DNA methylation data from infant cord blood has been generated at Scilifelab Uppsala (Responsible for analysis: Tomas Axelsson, [Tomas.Axelsson@medsci.uu.se](mailto:Tomas.Axelsson@medsci.uu.se)). This phase included data from 726 cord blood samples. Individuals were selected based in the following criteria:

- First time mother
- Mother donated blood during pregnancy
- Mother donated urine during pregnancy
- Cord blood available from delivery

### GWAS data

GWAS data from infant cord blood has been generated at Scilifelab Uppsala (Responsible for analysis: Susanne Björnefeldt, [snp@medsci.uu.se](mailto:snp@medsci.uu.se)). This phase will include data from the same 726 children described above.

### Chemical exposure data

Data on the following chemical compounds has been generated at the department of chemistry at Umeå University (contact Lisa Lundin, [lisa.lundin@umu.se](mailto:lisa.lundin@umu.se)).

- Parabens
- Brominated flame-retardants
- Polychlorinated biphenyls
- Organophosphates

## What types of data will be created and/or collected, in terms of data format and amount/volume of data?

### Questionnaire data

See description of collection above.

NorthPop is responsible for the collection, description and storage of QA data from their questionnaires. At request, specific variables will be created.

This data is classified as personal data and specific restrictions apply.

### DNA methylation data.

See description of collection above.

High dimensional data generated at scilifelab.

This data is classified as personal data and specific restrictions apply.

### GWAS data.

See description of collection above.

High dimensional data generated at scilifelab.

This data is classified as personal data and specific restrictions apply.

### Chemical exposure data

See description of collection above.

High dimensional data generated at the department of chemistry.

This data is classified as personal data and specific restrictions apply.

## Documentation and data quality

### How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?

Data will be divided into:

- RAW data files
  - Stored on a secure server at ITS and in a safe at the department of radiation sciences
  - Folder: ITS\_servername/Harlid\_NP\_VR/**Rawdata**/Datatype\_date of collection
  - **Kept as is**, for record keeping and to be able to backtrack
  - Original variable lists with NP QA data
  - Original IDAT files from Scilifelab, and quality control report
  - Original result-files from the department of chemistry
- Master files
  - Files with **merged data** containing QA variables, Methylation and chemical measurements
  - Folder: ITS\_servername/Harlid\_NP\_VR/**Masterfiles**/Date and version
  - Will be continuously updated, however older versions are retained for some time depending on storage space.
  - Statistical analyses are not performed on masterfiles
- Work files
  - Files that are in current use for different project analyses
  - May contain all data from the Masterfile or data subsets, depending on what research question is being investigated
  - All changes from the Masterfile should be documented in accompanying text file

- Folder: ITS\_servername/Harlid\_NP\_VR/**Workfiles**/Date and version
- May be “checked out” from the server, data and name of person using file must be documented in accompanying text file, file must be encrypted if stored on an individual computer.
- When analysis is complete the final version of the file must be stored in a separate folder: ITS\_servername/Harlid\_NP\_VR/**Workfiles/Completed\_projects**
- After analysis completion no versions can be kept on individual computers.

**How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?**

Quality control will be performed for individual datasets at Scilifelab and the department of chemistry, when the data is generated. QA-reports will be saved in files as described above.

For methylation data and GWAS data separate QA and normalization were performed upon data delivery.

## Storage and backup

**How is storage and backup of data and metadata safeguarded during the research process?**

All raw data files will be stored at a secure server at ITS (as described above). Individual work files, if they are >5GB can be stored on the project specific trygg filyta (<https://tfy.its.umu.se/grp/Harlid-VR-projekt>). However, this storage is only temporary and after the project is concluded all files will be stored at ITS only.

Metadata will be created at the end of the project and stored at NorthPop. Access to metadata will only be granted through NorthPop project administrators.

**How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?**

Since all data in this project (with the exception of Metadata described above) is considered sensitive and personal the safety measures preciously described does not differ depending on data types.

## Legal and ethical aspects

**How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?**

The data handling in NorthPop is described in the ethical permission (DNR: 2014/224-31).

Briefly: All data will be stored on secure locations only, such as a secure server at ITS, an external harddrive locked in safe at the department of radiation sciences or (temporarily) on “trygg filyta”. All files will be password protected and encrypted.

**How is correct data handling according to ethical aspects safeguarded?**

Ethical aspects in NorthPop is described in the ethical permission (DNR: 2014/224-31).

Results will be presented at group level only, individuals will not be identified.

## Accessibility and long-term storage

**How, when and where will research data or information about data (metadata) be made accessible? Are there any**

**conditions, embargoes and limitations on the access to and reuse of data to be considered?**

Research data will be made public through publication in scientific journals at the end of individual projects. Metadata will be accessible through NorthPop.

**In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?**

After project completion, data will be transferred to the NorthPop database for long term storage.

**Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or use/analyse data in the long term?**

This information is not yet available.

**How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?**

Personal identification numbers will be safeguarded as outlined above.

## **Responsibility and resources**

**Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?**

NorthPop: Richard Lundberg, [richard.lundberg@umu.se](mailto:richard.lundberg@umu.se)

Other project data management: Sophia Harlid, [sophia.harlid@umu.se](mailto:sophia.harlid@umu.se)

**What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?**

This information is not yet available.